# Recombinant expression and purification of an ATP-dependent DNA ligase from *Aliivibrio salmonicida*

Adele Williamson[1*], Hege Pedersen[2]

1. Department of Chemistry, University of Tromsø, N-9019 Tromsø Norway
2. Institute of Medical Biology, University of Tromsø, N-9019 Tromsø Norway

* Corresponding author
Phone: +47090637967
Email: adele.k.williamson@uit.no

## Abbreviations used

AMP, adenosine monophosphate; Amp, ampicillin; ATP, adenosine triphosphate; BLAST, Basic Local Alignment Search Tool; Cam, chloramphenicol ; CDS, coding DNA sequence; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; GST, glutathione-s-transferase; IPTG, isopropyl β-D-1-thiogalactopyranoside; LB, lysogeny broth; MBP, maltose-binding protein; NAD, nicotinamide adenine dinucleotide; Ni-IMAC, nickel-immobilized affinity chromatography; OD, optical density; RCF, relative centrifugal force; SDS-PAGE, sodium dodecyl sulfate- polyacrylamide gel electrophoresis; TBE, Tris/Borate/EDTA buffer; TEV protease, Tobacco Etch Virus protease

## Keywords

ATP-dependent DNA ligase; periplasmic localization sequence; *Aliivibrio salmonicida;* maltose-binding protein; recombinant expression

## Abstract

The genome of the psychrophilic fish-pathogen *Aliivibrio salmonicida* encodes a putative ATP-dependent DNA ligase in addition to a housekeeping NAD-dependent enzyme. In order to study the structure and activity of the ATP dependent ligase *in vitro* we have undertaken its recombinant production and purification from an *E. coli* based expression system.

Expression and purification of this protein presented two significant challenges. First, the gene product was moderately toxic to *E. coli* cells, second it was necessary to remove the large amounts of *E. coli* DNA present in bacterial lysates without contamination of the protein preparation by nucleases which might interfere with future assaying. The toxicity problem was overcome by fusion of the putative ligase to large solubility tags such as maltose-binding protein (MBP) or Glutathione-s-transferase (GST), and DNA was removed by treatment with a nuclease which could be inhibited by reducing agents.

As the *A. salmonicida* ATP-dependent DNA ligase gene encodes a predicted leader peptide, both the full-length and mature forms of the protein were produced. Both possessed ATP-dependent DNA ligase activity, but the truncated form was significantly more active. Here we detail the first reported production, purification and preliminary characterization of active *Aliivibrio salmonicida* ATP-dependent DNA ligase.

## Highlights

- A putative periplasmic DNA ligase from *Aliivibrio salmonicida* was recombinantly produced in *E. coli*
- Both the full-length and N-terminally truncated variants were expressed and purified
- Toxicity of gene products to the host cells was overcame by fusion to large solubility tags
- The truncated protein is more active in DNA ligation than the full-length version

## Introduction

DNA ligases are enzymes which catalyze the formation of a phosphodiester bond between adjacent 5’ $PO_4$ and 3’ OH ends in double-stranded DNA, and are essential for sealing breaks during DNA replication and repair [1]. DNA ligases can be divided into two types based on the nucleotide cofactor they use as an AMP donor: ATP-dependent DNA ligases (EC 6.5.1.1) which are found in eukaryotes and archaea, and NAD-dependent DNA ligases (EC 6.5.1.2) which are found exclusively in bacteria [2]. In addition to their house-keeping NAD-dependent enzymes, many bacteria have one or more ATP-dependent DNA ligases, the evolutionary origin and cellular function of which have not been entirely determined [3]. A number of these accessory enzymes have been biochemically characterized and some are postulated to play a role in DNA repair [4-7] while others are suggested to be involved in competence and DNA uptake [8, 9]. The genome of the pathogenic psychrophile *Aliivibrio salmonicida* encodes one such putative ATP-dependent DNA ligase [10].

In order to study its structure and activity *in vitro* we have undertaken the recombinant production and purification of this DNA ligase from an *E. coli* based expression system. Expression and purification of this protein presented two significant challenges. First, the gene product was moderately toxic to *E. coli* cells, second, crude lysates contained a large amount of bacterial DNA which needed to be removed prior to purification without contaminating the ligase protein preparation. Numerous publications have focused on the utility of large fusion partners in increasing protein solubility and expression levels (for example see [11, 12]), and comprehensive protocols for the production of MBP fusion constructs are available [13]. However the application of large fusion partners to overcome toxic effects of intracellularly-expressed proteins on the host cells has not been systematically reported to the same extent. In the case of the two Vib-Lig variants described here, the decreased host-cell growth rate with smaller tags presented a significant loss of efficiency during protein production, even before solubility issues were taken into consideration.

As the *A. salmonicida* gene encodes a predicted leader peptide, both the full-length and mature forms of the protein were produced, and ATP-dependent DNA ligation activity was verified for both constructs. This work represents the first instance of successful production, purification and preliminary characterization of active *Aliivibrio salmonicida* ATP-dependent DNA ligase.

# Methods

## Bioinformatics

The CDS YP_002262821.1 of the *A. salmonicida* genome encodes a 284 amino acid product VSAL_I1366 which is annotated as an ATP-dependent DNA ligase. BLAST homology searches show it has low identity with previously characterized DNA ligases from *Haemophilus influenza* (37%) *Neisseria meningitides* (36%) *Pseudomonas aeruginosa* (30%) and *Mycobacterium tuberculosis* homologues B, C and D (23% 24% 27%)[8, 9, 14, 15]. In spite of this low identity, VSAL_I1366 has a number of conserved residues which are involved in DNA ligase activity in homologues, including lysine 52 which is found in the motif I KxDG and is the site of AMP binding. Consistent with the gene annotation, a search of the pfam database [16] identifies two conserved domains: an N-terminal DNA-ligase adenylation domain (pfam01068) from residues 29 – 202 and a C-terminal oligonucleotide binding domain (cd08040) from residues 216 – 281.

Analysis with the program SignalP [17] using the 'gram negatives' network produced a strong prediction for a 21 amino acid leader peptide (mean S=0.726, meanD=0.692) with a cleavage position between residues Ala 21 and Phe 22 which would direct the enzyme to the periplasm of the

bacterial cell [18]. Analysis of the N-terminal sequence hydropathicity [19], along with helix-forming [20] and trans membrane tendencies [21] indicated that this predication falls within a hydrophobic helical sequence (Figure 1). For this reason, during cloning we truncated the polypeptide by a further four amino acids to what we believe is the beginning of the soluble functional domain. The 284 amino-acid full-length protein, denoted FL-Vib-Lig, has a computed molecular mass of 31.7 kDa and estimated pI of 5.68, while the N-terminally truncated protein, denoted TR-Vib-Lig, is 29.1 kDa and has a predicted pI of 5.51

**Image file: Fig_1_Leader_seq single column**

**Figure 1**. Sequence analysis of the N-terminal region of the putative ATP-dependent DNA ligase from *A. salmonicida.* The arrow indicates the cleavage position suggested by the 'SignalP' program. The bold underlined residues indicate the N-terminal sequence used below for cloning the truncated form.

## Cloning

The genes for FL-Vib-Lig and TR-Vib-Lig were both amplified in two stages using Phusion polymerase (New England BioLabs). In the first step, the coding sequence was amplified from a plasmid harboring the full-length *A. salmonicida* ATP DNA ligase gene using reverse primer BK 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAATATTTTTCACGAACC-3' and either the forward primer FD 5'-GAAAATCTTTATTTTCAAGGTAAAGTATCAACATTATCG-3'to produce the full-length sequence omitting the codon for the first methionine, or primer FD 5'-GAAAATCTTTATTTTCAAGGTAATACAGTCCCTGTTTCTGTATTG-3' to produce an N-terminal truncation lacking the first 25 amino acids corresponding the signal peptide. In the second step the PCR products were extended using the primers FD 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTC CATCACCATCACCATCACGAAAATCTTTATTTTCAAGGT-3' and the same reverse primer as in the first PCR reaction to add non-coding attB1 and attB2 sequences for recombination using the Gateway® system (Life Technologies), and coding sequences for an N-terminal 6-His tag followed by a TEV protease cleavage site [22]. These genes were used as substrates in the Gateway® BP reaction with pDONR221 to produce entry vectors containing the FL- or TR- genes. These entry constructs were transformed into One Shot® TOP10 Chemically Competent *E. coli* cells (Life Technologies), verified by sequencing with M13 universal primers (Sigma-Aldrich), and used as substrates for the Gateway®LR reaction with the destination vectors pDEST 17, pDEST 16, pDEST 15 (Invitrogen) and pHMGWA (GenBank #Eu680841) to produce four expression constructs for each gene (Table 1). As the PCR product included a coding sequence for an N-terminal 6-His tag and TEV cleavage site, all expression products were, in principle, able to be purified using nickel-immobilized affinity chromatography (Ni-IMAC) and the fusion tags removed by treatment with TEV protease. Because of this design, proteins

expressed from vectors pDEST 17 and PHMGWA had two His- tags. The products of the LR reaction were transformed into TOP10 cells for plasmid propagation, and subsequently into BL21 Star™ (DE3) pLysS *E. coli* (Life Technologies) for protein expression.

**Table 1.** Overview of construct design. His (6-Histine), TRX (thioredoxin) GST (Glutathione-s-transferase), MBP (maltose-binding protein). Polypeptide length and molecular weight were calculated from the coding regions of each fusion protein.

| Ligase variant | Protein name | Destination vector | Fusion tags | Length (amino acids) | Molecular weight (kDa) |
|---|---|---|---|---|---|
| FL | FL-Vib-Lig- His | pDEST17 | His-His | 318 | 35.9 |
| | FL-Vib-Lig- TRX | pDEST16 | TRX-His | 417 | 46.4 |
| | FL-Vib-Lig- GST | pDEST15 | GST-His | 535 | 61.1 |
| | FL-Vib-Lig- MBP | PHMGWA | His-MBP-His | 690 | 76.4 |
| TR | TR-Vib-Lig- His | pDEST17 | His-His | 293 | 33.3 |
| | TR-Vib-Lig- TRX | pDEST16 | TRX-His | 392 | 43.8 |
| | TR-Vib-Lig- GST | pDEST15 | GST-His | 510 | 58.5 |
| | TR-Vib-Lig- MBP | PHMGWA | His-MBP-His | 665 | 73.8 |

## Expression testing

Expression of recombinant tagged protein was investigated at either 37 $^{o}$C or 22 $^{o}$C in small-scale cultures. Freshly transformed BL21 star pLysS pRARE cells were grown overnight at 37 $^{o}$C with mixing on an oribital shaker, and 25 µl of this dense pre-culture was used to inoculate 5 mLs of LB medium containing 50 µg mL$^{-1}$ ampicillin and 34 µg mL$^{-1}$ chloramphenicol (LB/amp/cam). Cultures were grown at 37 $^{o}$C for 3 hours after which time protein expression was induced by addition of 1mM IPTG, and the temperature was either maintained at 37 $^{o}$C for a further 9 hours, or decreased to 22 $^{o}$C and incubated overnight before harvesting. Cell density was monitored by recording the $OD_{600}$ every hour, and after the incubation time had elapsed, cells were harvested by centrifugation at 4000 RCF and cell pellets frozen at -20 $^{o}$C. Cell lysis for verification of protein expression was done by resuspending the cell pellet in 200 µl of 'Fastbreak' buffer (Promega) per 5 mLs culture with 0.2mg mL$^{-1}$ lysozyme and 0.02 mg mL$^{-1}$ DNAseI (Sigma-Aldrich D2821). Samples for solubility screening were harvested after 4 hours (37 $^{o}$C) or overnight (22 $^{o}$C) and resuspended in 1 mL per 5 mLs of culture in 50 mM Tris pH 8.0, 200 mM NaCl, 1mM DTT, 10% glycerol, 2mg mL$^{-1}$ lysozyme and 0.02mg mL$^{-1}$ DNAse, followed by mechanical lysis using a Q125 sonicator (Qsonia) with the microtip

set at 20% intensity, in 1s pulses on/off for a total of 1 minute on ice. Insoluble material was pelleted by centrifugation at 17,000 RCF for 10 minutes at 4 $^{o}$C.

Samples were electrophoresed on 4–20% Mini-PROTEAN® TGX™ precast polyacrylamide gels (Bio-Rad) and protein bands visualized by staining with SimplyBlue™ SafeStain (Life technologies) according to the manufacturers' instructions.

## Large scale purification

Based on the results of the expression and solubility testing, the His-MBP-His-tagged constructs were chosen for large scale recombinant production of both Vib-Lig variants. Expression cultures were inoculated with 30 mLs of pre-culture per L of LB/amp/cam, and grown at 37 $^{o}$C with vigorous shaking to an $OD_{600}$ of 0.3 – 0.4, which typically took 2 – 3 hours. After this time the temperature was decreased to 22 $^{o}$C and cultures were allowed to equilibrate for 30 minutes before induction of protein expression by addition of 1mM IPTG. Expression was continued overnight before harvesting and storage as above.

Lysis and all subsequent purification steps were carried out at 4 $^{o}$C. Thawed cells were resuspended in lysis buffer (50mM Tris pH 8.0, 750 mM NaCl, 1 mM $MgCl_2$, 1mM DTT 0.1 mM ATP, 5% glycerol) at 5 mLs per g of cell pellet and lysed by two passages through a French press (Constant Systems) at 18 PIS. Insoluble material was removed by centrifugation at 17,000 RCF for 35 minutes and the crude soluble fraction was incubated with the non-specific nuclease 'HL-SAN' at 1 unit per mg cell pellet overnight at 4 $^{o}$C. HL-SAN is a heat-labile mutant of the commercially available 'Salt Active Nuclease' (SAN) which our group is beta-testing and was not on the market at the time of writing; for enquiries contact the company (Arcticzymes Norway, www.arcticzymes.com). The treated lysates were loaded onto 5 mL His Traps columns (GE Healthcare) equilibrated with Buffer A (50mM Tris pH 8.0, 750 mM NaCl, 5% glycerol, 10 mM imidazole) and washed with 25-50 mLs of Buffer A before elution on a 15 mL gradient from 0 – 100 % Buffer B (50mM Tris pH 8.0, 750 mM NaCl, 5% glycerol, 500 mM imidazole). The eluted fractions were exchanged into Buffer C (50mM Tris pH 8.0, 200 mM NaCl, 1mM DTT, 5% glycerol) using a HiPrep 26/10 column and incubated overnight with TEV protease (produced in house according to [23]) at 0.1 mg TEV per 1ml of eluted fusion protein. Protease-treated protein was subjected to a reverse Ni-IMAC step using a 5ml His Trap equilibrated with Buffer C. The flow-through was collected, up concentrated to approximately $A_{280}$ 2.0 and loaded onto a Superdex200 16/600 column at 1 mL/min. Peaks eluting from the column were analyzed by SDS-PAGE as described above and verified by quadrupole-time of flight liquid chromatography–mass spectrometry (Q-ToF LC-MS/MS). Pure Vib-Lig protein was mixed 50:50 v/v with glycerol and stored at -80 $^{o}$C.

## Assay for ligase

DNA ligase activity was verified using an assay similar to that described in [24]. An 18 nt nicked double-stranded DNA substrate where one of the 9-mer oligonucleotides on the nicked strand was fluorescently labeled with the FAM moiety (Sigma-Aldrich). FL-Vib-Lig or TR-Vib-Lig were incubated at concentrations indicated in the figure legends with 80 nM substrate, 0.1 mM ATP, 10 mM MgCl2, 1mM DTT for 10 minutes at 15 $^{o}$C. The reaction was quenched by addition of 25% formamide, 2 mM EDTA and heating to 95 $^{o}$C for 5 min. Denaturing electrophoresis was carried out subsequent to the reaction on a 20% acrylamide 7M urea 1x TBE gel, and the FAM-labeled oligonucleotides were visualized on a Pharos FX Plus imager (Biorad). The intensity of bands corresponding to the ligated (18 nt) product and un-ligated (9 nt) substrate were integrated using the software Image J [25], and the extent of ligation was taken as the ratio of the two bands, expressed as a percentage.

# Results and discussion

## Cloning and expression

Both the full-length and N-terminally truncated forms of the *vib-lig* gene were successfully cloned into the entry vector pDONR 221 and confirmed by Sanger sequencing.

A comparison of small-scale cultures expressing either of the Vib-Lig variants at 37 $^{o}$C from the four pDEST vectors used in this study shows a strong correlation between the tag size and growth rate (Figure 2). Smaller tags, His-His (4.2 kDa) and TRX-His (14.7 kDa), had significantly slower rates of growth relative to fusion constructs with larger proteins, GST-His (29.4 kDa) and His-MBP-His (44.7 kDa). Similar results were obtained for cultures grown at 22 $^{o}$C overnight after induction.

Examination of total crude lysate from expression cultures confirmed the presence of highly expressed protein at the expected molecular weights for each construct of each Vib-Lig variant (Figure 3). Non-induced controls all had similar growth rates regardless of the construct they harbored, and little or no leaky protein production was detected by SDS-PAGE (data not shown). There was little difference between growth rates for the FL- and TR- variants of Vib-Lig with exception of Vib-Lig-TR-MBP which grew poorly for four hours after induction, but reached ODs comparable to that of other large-tagged constructs by the end of the time course.

Taken together these results suggest that the Vib-Lig protein is moderately toxic to the *E. coli* cells when overexpressed at either 37 or 22 $^{o}$C; however this toxicity can be overcome by fusion to proteins of similar or larger size. Although a decrease in growth rate following induction of protein expression can be expected due to the metabolic burden on cells, this is not likely to be the case

here as GST- and MBP- tagged constructs approached the growth rates of non-induced cultures, while the level of protein overexpression was similar for all constructs. We suggest two mechanisms that could produce these toxic effects on the expression host: binding of intracellular ATP which perturbs the energy balance inside the cell, or binding of DNA which interferes with replication and/ or protein expression. In both cases, steric hindrance by a bulky protein fusion partner could prevent this behavior- either by blocking the DNA binding site of the ligase protein directly, or by hindering the domain re-arrangement which accompanies adenylation and DNA binding in all homologous studied to date [26]. These hypotheses could be confirmed by examining the DNA- and ATP- binding affinities of purified Vib-Lig variants with the different tags; however our primary objective of producing large amounts of recombinant Vib-Lig protein has been met and thus these phenomena were not investigated further.

Solubility screening of crude lysates showed that His-His- or TRX-His constructs were predominantly insoluble, while GST-His- tagged fusions had soluble product at both temperatures (Table 2). MBP-tagged constructs of both variants were somewhat soluble at 37 $^{o}$C, and highly soluble at 22 $^{o}$C. This is consistent with the work by Niiranen et. al which surveyed a range of fusion partners for the expression of *A. salmonicida* proteins, and showed that fusion to MBP, coupled with a decreased cultivation temperature to be one of the most useful in increasing protein solubility [27]. It is interesting to note that GST-Vib-Lig fusions were markedly more soluble compared to the smaller tagged constructs, despite GST being a poor solubility enhancer in previous studies [28] [27].

Due to the superior growth rates of expression cultures and increased solubility of protein products, large scale production of both Vib-Lig variants was carried out using the MBP-tagged constructs at 22 $^{o}$C.

**Image file: Fig_2_Growth double column**

**Figure 2.** Growth at 37 $^{o}$C of *E. coli* BL32(DE3)pLysS star pRARE expressing Vib-Lig variants as fusion proteins with tags as described in the text: Growth profiles over 12 hour time course, A) Full-length (FL) Vib-Lig and B) Truncated (TR) Vib-Lig. C) Growth rates of cultures during the 4 hours following induction. Rates were derived from a linear fit of points between 3 and 7 hours. Graph values indicate the mean of three replicates; error bars indicate standard deviation of each triplicate.

**Image file: Fig_3_expression single column**

**Figure 3.** SDS-PAGE gel of total crude lysate from *E. coli* BL32(DE3) pLysS star pRARE expression cultures of Full-length Vib-Lig (FL) and truncated Vib-Lig (TR) with His, TRX, GST and MBP fusion tags as described in the text. Markers indicate the approximate band position expected for each fusion protein; see text for details.

**Table 2.** Solubility screening of recombinantly expressed Full-length (FL)- and truncated (TR) Vib-Lig variants with His, TRX, GST and MBP fusion tags as described in text. Scores are based on the presence of a band of the correct molecular weight in soluble crude fractions analyzed by coomassie-stained SDS-PAGE: ++ strong band, + clearly visible, (+) weak, - absent.

| Variant | Construct | 37 °C | 22 °C |
|---|---|---|---|
| FL | Vib-Lig-FL-His | - | - |
| | Vib-Lig-FL-TRX | - | - |
| | Vib-Lig-FL-GST | + | + |
| | Vib-Lig-FL-MBP | + | ++ |
| TR | Vib-Lig-TR-His | - | (+) |
| | Vib-Lig-TR-TRX | - | - |
| | Vib-Lig-TR-GST | + | + |
| | Vib-Lig-TR-MBP | + | ++ |

## Large scale expression and purification

Large-scale cultivation of both MBP-tagged FL- and TR- constructs using the conditions described above produced 4- 5 g $L^{-1}$ of cell pellet. After breakage by French press, the crude cell lysate was extremely viscous and difficult to load onto the chromatography columns; hence removal of contaminating DNA was necessary. We found that extended sonication produced variable results and often increased protein precipitation, while treatment of the lysate with bovine DNAseI lead to detectable DNAse activity in some batches of 'pure' protein (data not shown). However, treatment with HL-SAN successfully removed high molecular weight DNA form the sample (Figure 4) without causing any residual DNAse contamination (data not shown).

**Image file: Fig_4_DNase single column**

**Figure 4.** The effect of HL-SAN treatment on high-molecular weight DNA content of crude cell lysates: lane 1. 1 Kb DNA ladder, 2. Untreated 3. HL-SAN treated

The first IMAC step in the purification of both Vib-Lig variants yielded a single peak at around 80 % Buffer B (data not shown). SDS-PAGE of both variants showed major protein bands between 55 and 66 kDa which correspond to the molecular weights of Vib-Lig- MBP- fusion proteins and the *E. coli* chaperonin protein GroEL 60; in the case of FL-Vib-Lig this assignment was confirmed by MS/MS analysis (Table 3). Cleavage with TEV protease produced bands of the size expected for the tag-free Vib-Lig variants, and the MBP-His tag; the former was confirmed by MS/MS. A second reverse Ni-IMAC step removed virtually all MBP-His cleavage product, leaving a semi-crude preparation containing tag-free Vib-Lig, along with a small amount of fusion protein and *E. coli* chaperonin. In the

case of TR-Vib-Lig these contaminants were successfully removed in a single step of gel filtration where a peak eluting with a retention time of approximately 40 mLs contained chaperonin and fusion protein, while a later fraction was highly pure TR-Vib-Lig. Separation of FL-Vib-lig was more complicated as protein eluted from the gel filtration column in three major peaks corresponding to the column void volume and containing primarily un-cleaved FL-Vib-Lig- MBP (Figure 5 A, lane 5), a complex of approximately 800 kDa comprising primarily chaperonin (Figure 5 A, lane 6), and a species of 30 – 40 kDa (Figure 5 A, lane 7). All three fractions contained FL-Vib-Lig, with the most pure being the third peak. Our inability to remove the chaperonin contaminant during gel filtration was the primary reason that the yield of pure FL-Vib-lig was so low, less than 0.1 mg per L of culture (0.015 mg per g cells), compared to that of TR-Vib-Lig which was 1.8 – 2.0 mg L-1 (0.3 – 0.4 mg per g) of cells.

**Image file: Fig_5_purification single column**

**Figure 5.** Purification of FL-Vib-Lig (A) and Tr-Vib-Lig (B). Lane 1. Crude soluble lysate; 2. Fusion protein eluted from initial His-trap; 3. Cleavage of fusion protein with TEV-protease; 4. Flow-through from reverse IMAC containing tag-free Vib-Lig; 5, 6 and 7 fractions from gel filtration. The expected positions of bands corresponding to Vib-Lig-MBP fusion proteins, *E. coli* chaperonins, MBP-tag and tag-free Vib-Lig are indicated by arrows. Bands marked with an asterisk (*) were analyzed by mass spectrometry, and numbers correspond to results in Table 3.

**Table 3.** Results of MS/MS mass spectrometry from bands indicated (*) in Figure 5 searched against the NCBI database. Mass: mass of the matched peptide sequence.

| Band | Protein hit | Acession number | Mass | Matches | Sequences |
|---|---|---|---|---|---|
| 1. | maltose binding protein-lacZ | gi\|2623823 | 53069 | 58 | 13 |
| | DNA ligase [Aliivibrio salmonicida LFI1238] | gi\|209694893 | 31696 | 33 | 13 |
| 2. | *E. coli* Chaperonin Complex GroelGROES | gi\|2624772 | 57162 | 91 | 18 |
| 3. | DNA ligase [Aliivibrio salmonicida LFI1238] | gi\|209694893 | 31696 | 26 | 13 |
| 4. | DNA ligase [Aliivibrio salmonicida LFI1238] | gi\|209694893 | 31696 | 27 | 11 |

The difference in these elution profiles suggests that the *E. coli* chaperonin protein is more tightly bound to the FL-Vib-Lig variant; most likely as the chaperonin recognizes the hydrophobic N-terminal segment as a mis-folded or incorrectly-translocated protein. We also observed that FL-Vib-Lig was more challenging to manipulate than the truncated variant: it precipitated more readily and could not be up concentrated above 0.5 mg $mL^{-1}$.

## DNA ligase assay

Both Vib-Lig variants were able to ligate a phosphorylated nick in a double-stranded DNA substrate in the presence of ATP and $MgCl_2$ as shown by the appearance of the 18 nt band corresponding to the expected size of the ligation product (Figure 6 A). No ligated product was detected in the control reaction, confirming that the appearance of the upper band is strictly dependent on the presence of the enzyme. Dosage response experiments (Figure 6 B), and progress curves (Figure 6 C) reveal that TR-Vib-Lig is significantly more active than the full-length variant. This indicates that the presence of the N-terminal sequence interferes with enzymatic activity, in addition to protein stability, and strongly supports the assignment of these residues as part of a localizing sequence rather than a feature of the native mature enzyme. The dosage response also reveals that even at saturating concentrations of TR-Vib-Lig only 80% of the substrate is ligated, which probably reflects incomplete annealing of the labeled oligonucleotide with its complementary strand. Experiments investigating cofactor preference show that omission of ATP from the reaction greatly reduces the activity of both forms of the protein, and addition of NAD cannot be substituted (Figure 6 D). This confirms the bioinformatic assignment of this protein as an ATP-dependent enzyme. The residual ATP-independent activity is probably due to a significant fraction of the enzyme being pre-adenylated from addition of ATP to the lysis buffer during purification. Many previously described ATP-dependent DNA ligases were purified as the enzyme-adenylate [6, 15, 29] with this intermediate comprising up to 90% of the pure protein when measured by mass spectrometry [9]. Addition of EDTA abolishes ligase activity in both forms of the protein, while addition of $MnCl_2$ in place of $Mg\ Cl_2$ can support some activity, indicating that a divalent metal ion is essential as is the case for all previously characterized homologues.

**Image file: Fig_6_Assay double column**

**Figure 6.** DNA ligase activity of FL-Vib-Lig and TR-Vib-Lig measured by nick-sealing of a double-stranded DNA substrate. Standard reaction conditions were as described in the text unless otherwise stated. Graph points indicate the mean of three replicates; error bars indicate standard deviation of each triplicate. A representative gel corresponding to the plotted data is shown for each experiment. (A) Conversion of the 9 nt substrate to 18 nt product in the presence of either FL-Vib-Lig (75 nm ) or Tr-Vib-Lig (78 nM), but not in the absence of enzyme. The marker is a mixture of FAM-labeled oligonucleotides with size indicated in the figure. (B) Extent of ligation as a function of the concentration of FL-Vib-Lig (triangles) and TR-Vib-Lig (circles). (C) Time course of FL-Vib-Lig (triangles) and TR-Vib-Lig (circles) ligase activity. Concentrations were 15.5 nm and 15.0 nM respectively. (D) The effect of different cofactors on ligase activity of FL-Vib-Lig (left) and TR-Vib-Lig (right). Values are expressed as a percentage of the TR-Vib-Lig control reaction. The following modifications to the standard reaction conditions were used: ATP excluded (No ATP); ATP replaced by 26 μM NAD ($NAD^+$); $MgCl_2$ excluded, 5mM EDTA added (EDTA); $MgCl_2$ excluded, 5 mM $MnCl_2$ added. Enzyme concentrations were 78 nM and 75 nM respectively.

The results presented here strongly suggest that TR-Vib-Lig represents the mature, biologically relevant form of the VSAL_I1366 protein, thus it is this variant that should be used in future biochemical and structural studies. Two previously characterized ATP-dependent DNA ligases from

*Haemophilus influenza* and *Neisseria meningitides* also possess putative periplasmic leader sequences. In both cases these proteins were produced only as full-length constructs [8, 9], hence our comparison of FL-Vib-Lig and TR-Vib_lig constitutes the first, albeit preliminary, biochemical investigation of the significance of the leader peptide. Magnet et. al propose a biological function for this enzyme in natural competence of the host bacteria, and we hope that further *in vivo* and *in vitro* research on the Vib-Lig protein will allow detailed testing of such hypotheses.

## Conclusion

In conclusion, we have recombinantly expressed and purified the product of the YP_002262821.1 gene from *A. salmonicida* and confirmed that it possesses ATP-dependent DNA ligase activity. Comparison of the full-length and truncated versions of the protein indicate that the N-terminal 25 amino acids encode a polypeptide which does not form a functional domain of the enzyme, and based on bioinformatic analysis is most likely a periplasmic localization sequence in the native bacterium. This protein preparation will serve as a basis for more detailed characterization of the enzymatic and structural features of this protein in the future, which will help us to understand its biological function.

## Acknowledgments

We would like to thank Prof. Arne Smalås and Dr Elin Moe for their constructive comments and the Tromsø University Proteomics Platform (TUPP) for technical assistance in protein identification by mass spectrometry.

This work was supported by the Research Council of Norway (grant no. 192123).

## References


[1] D.J. Timson, M.R. Singleton, D.B. Wigley, DNA ligases in the repair and replication of DNA. Mutation Research-DNA Repair 460 (2000) 301-318.

[2] A.E. Tomkinson, S. Vijayakumar, J.M. Pascal, T. Ellenberger, DNA ligases: structure, reaction mechanism, and function. Chem Rev 106 (2006) 687-699.

[3] A. Wilkinson, J. Day, R. Bowater, Bacterial DNA ligases. Mol Microbiol 40 (2001) 1241-1248.

[4] C.L. Gong, P. Bongiorno, A. Martins, N.C. Stephanou, H. Zhu, S. Shuman, M.S. Glickman, Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. Nature Structural & Molecular Biology 12 (2005) 304-312.

[5] R.S. Pitcher, A.J. Green, A. Brzostek, M. Korycka-Machala, J. Dziadek, A.J. Doherty, NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. DNA Repair 6 (2007) 1271-1276.

[6] H. Zhu, S. Shuman, Characterization of Agrobacterium tumefaciens DNA ligases C and D. Nucleic Acids Research 35 (2007) 3631-3645.

[7] M. Blasius, R. Buob, I.V. Shevelev, U. Hubscher, Enzymes involved in DNA ligation and end-healing in the radioresistant bacterium Deinococcus radiodurans. Bmc Molecular Biology 8 (2007).

[8] C.H. Cheng, S. Shuman, Characterization of an ATP-dependent DNA ligase encoded by Haemophilus influenzae. Nucleic Acids Research 25 (1997) 1369-1374.

[9] S. Magnet, J.S. Blanchard, Mechanistic and kinetic study of the ATP-dependent DNA ligase of Neisseria meningitidis. Biochemistry-Us 43 (2004) 710-717.

[10] E. Hjerde, M.S. Lorentzen, M.T.G. Holden, K. Seeger, S. Paulsen, N. Bason, C. Churcher, D. Harris, H. Norbertczak, M.A. Quail, S. Sanders, S. Thurston, J. Parkhill, N.P. Willassen, N.R. Thomson, The genome sequence of the fish pathogen Aliivibrio salmonicida strain LFI1238 shows extensive evidence of gene decay. Bmc Genomics 9 (2008).

[11] K.D. Pryor, B. Leiting, High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system. Protein Expr Purif 10 (1997) 309-319.

[12] S. Nallamsetty, B.P. Austin, K.J. Penrose, D.S. Waugh, Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli. Protein Sci 14 (2005) 2964-2971.

[13] P. Sun, J.E. Tropea, D.S. Waugh, Enhancing the solubility of recombinant proteins in Escherichia coli by using hexahistidine-tagged maltose-binding protein as a fusion partner. Methods in molecular biology (Clifton, NJ) 705 (2011) 259-274.

[14] H. Zhu, S. Shuman, A primer-dependent polymerase function of pseudomonas aeruginosa ATP-dependent DNA ligase (LigD). Journal Of Biological Chemistry 280 (2005) 418-427.

[15] C.L. Gong, A. Martins, P. Bongiorno, M. Glickman, S. Shuman, Biochemical and genetic analysis of the four DNA ligases of mycobacteria. Journal Of Biological Chemistry 279 (2004) 20594-20606.

[16] M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. Finn, The Pfam protein families database. Nucleic Acids Research 40 (2012) D290-D301.

[17] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth 8 (2011) 785-786.

[18] G. von Heijne, The signal peptide. The Journal of membrane biology 115 (1990) 195-201.

[19] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein. J Mol Biol 157 (1982) 105-132.

[20] P.Y. Chou, G.D. Fasman, Prediction of the secondary structure of proteins from their amino acid sequence. Advances in enzymology and related areas of molecular biology 47 (1978) 45-148.

[21] G. Zhao, E. London, An amino acid “transmembrane tendency” scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. Protein Sci 15 (2006) 1987-2001.

[22] D.A. Polayes, T.D. Parks, S.A. Johnston, W.G. Dougherty, Application of TEV Protease in Protein Production #, T Molecular Diagnosis of Infectious Diseases, 1997, pp. 169-183.

[23] J.E. Tropea, S. Cherry, D.S. Waugh, Expression and Purification of Soluble His6-Tagged TEV Protease, 2009, pp. 297-307.

[24] Y. Wang, J.J. Xie, Z. Han, J.H. Liu, X.P. Liu, Expression, purification and biochemical characterization of Methanocaldococcus jannaschii DNA ligase. Protein Expres Purif 87 (2013) 79-86.

[25] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. Nat Meth 9 (2012) 671-675.

[26] S. Shuman, DNA Ligases: Progress and Prospects. Journal of Biological Chemistry 284 (2009) 17365-17369.

[27] L. Niiranen, S. Espelid, C.R. Karlsen, M. Mustonen, S.M. Paulsen, P. Heikinheimo, N.P. Willassen, Comparative expression study to increase the solubility of cold adapted Vibrio proteins in Escherichia coli. Protein Expr Purif 52 (2007) 210-218.

[28] R.B. Kapust, D.S. Waugh, Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8 (1999) 1668-1674.

[29] N. Keppetipola, S. Shuman, Characterization of a thermophilic ATP-dependent DNA ligase from the euryarchaeon Pyrococcus horikoshii. J Bacteriol 187 (2005) 6902-6908.